

*J. Math. and Its Appl.*

ISSN: 1829-605X

Vol. 7, No. 2, November 2010, 13-25

PENGLASTERAN PASIEN KANKER LEUKEMIA BERDASARKAN DATA EKSPRESI GEN DENGAN MENGUNAKAN DEKOMPOSISI NILAI SINGULAR

Evi Noviani¹, Yoga Satria Putra²

¹Jurusan Matematika, ²Jurusan Fisika

FMIPA Universitas Tanjungpura Kalimantan

¹evi_noviani@mipa.untan.ac.id

Abstrak

Informasi yang terkandung di dalam rangkaian molekular Deoxyribonucleic acid (DNA) makhluk hidup dapat diketahui melalui teknologi microarray. Data microarray menyajikan data tingkat ekspresi gen yang umumnya berukuran besar. Satu sampel pada data microarray bisa memiliki ribuan atau puluhan ribu gen. Pada penelitian ini diolah data pasien kanker darah (leukemia) yang berukuran 500032 dengan entri tak negatif. Data microarray pasien leukemia dapat diolah dengan menggunakan Dekomposisi Nilai Singular sedemikian sehingga sampel yang memiliki sifat yang sama dikelompokkan dalam satu kelompok. Dekomposisi Nilai Singular digunakan untuk mengelompokkan data dengan dua macam keragaman (bi-clustering), yaitu menggunakan nilai vektor singular kedua dan ketiga. Dari implementasi pada data, pasien kanker dapat dikelompokkan menjadi penyakit AML, dan ALL beserta sub tipe penyakit ALL, yakni ALL-T dan ALL-B.

Katakunci: *Data microarray, leukemia, bi-clustering, dekomposisi nilai singular.*

1. Pendahuluan

Setiap makhluk hidup, termasuk manusia, tersusun atas unit kerja terkecil yang dinamakan sel. Setiap sel yang menyusun tubuh makhluk hidup melakukan berbagai macam pekerjaan seperti metabolisme, memperbanyak diri dan lain-lain, sesuai fungsi selnya secara teratur, sehingga makhluk hidup tersebut dapat melakukan berbagai macam aktifitas. Berbagai aktifitas yang dilakukan sel telah terkodekan dalam rangkaian kode informasi genetik yang terdapat dalam rantai panjang *Deoxyribonucleic acid* (DNA). Kode informasi genetik ini mengandung bagaimana ciri-ciri, sifat, dan semua informasi lengkap tentang makhluk hidup tersebut. Pada organisme eukariotik, organisme yang memiliki membran inti sel, rangkaian DNA ini terdapat dalam inti sel.

Rangkaian informasi genetik dalam DNA pada akhirnya merupakan instruksi untuk membentuk protein tertentu. Dari perkembangan ilmu biologi dalam bidang molekuler yang semakin pesat, terdapat teknologi *microarray* yang dapat menyajikan data tingkat ekspresi gen yang dapat mengukur proses pembentukan protein ini. Satu sampel pada *microarray* bisa memiliki ribuan atau ratusan ribu gen. Sampel ini bisa berupa banyaknya orang atau organisme yang diteliti pada jaringan tertentu, atau bisa juga dari data seorang atau satu organisme saja yang diambil pada waktu yang berbeda. Sehingga data ekspresi gen yang didapatkan berupa matriks ukuran besar dengan entri dari matriks yang menunjukkan level ekspresi dari gen pada suatu sampel.

Dari data *microarray* ini perlu dianalisis sehingga dapat diketahui informasi yang terkandung di dalamnya. Salah satu informasi yang bisa didapatkan adalah pengelompokan gen atau sampel. Sampel yang mempunyai sifat yang hampir sama digolongkan dalam kelompok yang sama dan yang memiliki sifat berbeda dalam kelas yang berbeda. Proses pengelompokan ini dikenal dengan nama pengklasteran. Pengelompokan tipe-tipe kanker pada pasien sangat diperlukan sebagai pendukung diagnosa sehingga penangananannya menjadi tepat. Terdapat jenis kanker dengan ciri fisik yang sama tetapi jenisnya berbeda. Jika terdapat kekeliruan dalam diagnosa maka akan memperparah kondisi pasien. Salah satu metoda yang dapat dipakai untuk mengklasterkan data adalah menggunakan dekomposisi nilai singular yang merupakan metoda yang dipakai pada penelitian ini.

2. Landasan Teori

Deoxyribonucleic acid (DNA) merupakan pembawa informasi utama pada sebuah sel [1]. Rangkaian instruksi yang tersimpan dalam DNA mengatur aktifitas sel seperti metabolisme dan pembelahan sel. DNA ini tersusun dari molekul kecil bernama nukleotida. Terdapat empat nukleotida berbeda yang digolongkan ke dalam dua grup, yakni *purin*, terdiri dari *adenosine* dan *guanine*, serta *pyrimidin* terdiri dari *cytosine* dan *thymine*. Keempat nukleotida ini dinotasikan secara berurutan dengan *A, G, C* dan *T*. Kemudian nukleotida ini membentuk rangkaian DNA yang panjang dan membentuk struktur *double helix*.

Rangkaian DNA ini memiliki kode-kode yang merepresentasikan ciri fisik makhluk hidup, cara bekerja sel dan lain-lain. Kerja sel dapat dilihat dari protein yang dihasilkan dari kombinasi DNA ini. DNA akan mengalami transkripsi menjadi RNA (*Ribo-nucleic acid*), yaitu rangkaian nukleotida seperti halnya DNA, namun nukleotida *Thymine* diganti dengan *Uracil* (U). Pada proses transkripsi, molekul DNA pada salah satu benang disalin menjadi pre mRNA. Proses selanjutnya adalah *splicing*, yaitu membuang potongan rangkaian DNA *introns* dan menyambung rangkaian yang akan dikodekan yakni *exons* menjadi satu rangkaian mRNA. Selanjutnya dilakukan proses translasi, mRNA menuju ke ribosom yang merupakan tempat translasi. mRNA ini terbaca oleh ribosom sebagai triplet codons. Rangkaian tiga nukleotida pada DNA mengkodekan asam amino. Pada proses translasi dibentuk protein dengan menyatukan asam amino yang terkodekan pada mRNA. Protein yang dihasilkan inilah yang dapat melaksanakan fungsi yang dibutuhkan oleh sel tubuh.

Central dogma of molecular biology menyebutkan bahwa informasi tidak dapat ditransfer kembali dari protein kepada protein yang lain atau kepada asam amino. Hal ini berarti satu rangkaian DNA akan mengkodekan ekspresi gen yang menghasilkan suatu jenis protein yang unik.

Seiring perkembangan teknologi, telah ditemukan metode untuk menemukan urutan rangkaian DNA secara lengkap pada suatu makhluk hidup. Selanjutnya, perkembangan teknologi juga berhasil menemukan teknologi cDNA dan *microarray* yang dapat mengukur tingkat ekspresi gen dari puluhan ribu gen secara parallel dan menghasilkan data yang sangat besar dan berharga [2]. Ekspresi gen muncul ketika informasi molekuler tertentu

yang mengandung DNA direkam untuk menampilkan RNA-nya. Molekul inilah yang membuat protein untuk melakukan berbagai fungsi sel. Menurut *Central dogma of molecular biology*, setiap gen akan menghasilkan satu mRNA yang kemudian menghasilkan satu protein, maka level ekspresi gen satu gen dengan gen yang lain akan berbeda-beda. Teknologi *microarray* dapat mengukur ekspresi gen dalam jumlah besar secara simultan. Fragmen DNA ditempatkan dalam slide, lalu RNA yang telah diekstrak dari jaringan dan diberi label *fluorescent*, akan menyebar dalam *microarray*. Selanjutnya mengubah scan data image menjadi bentuk ekspresi gen. Seperti yang diilustrasikan [3], contoh membuat data *microarray* dari organisme ragi (*yeast*) dan [6] yang menjelaskan prosedur pembuatan data *microarray*.

Data mentah pada eksperimen *microarray* berbentuk gambar. Untuk mendapatkan data mengenai tingkat ekspresi gen, gambar tersebut harus dianalisa. Masing-masing titik diidentifikasi, diukur intensitasnya dan dibandingkan dengan latar belakangnya. Data tersebut kemudian dimasukkan kedalam sebuah hasil akhir berupa matriks ekspresi gen. Matriks inilah yang nantinya dianalisa untuk proses lebih lanjut. Pada matriks ini, baris merepresentasikan gen, dan kolom merepresentasikan berbagai macam sampel. Angka pada baris dan kolom yang bersesuaian merepresentasikan tingkat ekspresi gen tertentu pada sampel tertentu.

Data ekspresi gen disajikan dalam matriks $W_{M \times N}$ dengan N merupakan banyaknya sampel, sedangkan M adalah banyaknya gen yang diteliti. Entri dari matriks W yaitu W_{ij} merupakan level ekspresi dari gen i pada sampel j . Nilai W_{ij} yang lebih besar diantara yang lain mengandung arti level aktifitasnya lebih besar dari yang lain.

Pada penelitian ini digunakan metode Dekomposisi Nilai Singular.

Teorema 2.1 (*Dekomposisi nilai singular*) Misalkan $A \in \mathbb{R}^{(m \times n)}$ maka terdapat matriks orthogonal $U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m}$, $V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$ dan matriks diagonal $\Sigma_1 \in \mathbb{R}^{r \times r}$ dengan $r \leq \min\{m, n\}$, sedemikian sehingga

$$A = U \Sigma V^T$$

Dengan Σ berukuran $m \times n$ yang mempunyai bentuk

$$\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$$

Vektor-vektor kolom pada U disebut vektor singular kiri dan vektor-vektor kolom pada V disebut vektor singular kanan.

Pada penelitian ini dikelompokkan data pasien kanker darah (leukemia) [4]. Leukemia adalah penyakit yang ditandai oleh adanya akumulasi leukosit ganas dalam sumsum tulang dan darah. Sel-sel abnormal ini menyebabkan timbulnya gejala karena :

Kegagalan sumsum tulang (anemia, netropenia, trombositopenia) dan Infiltrasi organ (hati, limpa, kelenjar getah bening).

Secara tradisional Leukemia diklasifikasikan berdasarkan tipe sel yang terlihat dan bentuk maturasi sel Leukemia, yaitu:

a. Leukemia akut [8]

Leukemia akut merupakan suatu penyakit yang serius, berkembang dengan cepat, dan apabila tidak diterapi dapat menyebabkan kematian dalam beberapa minggu atau bulan. Leukemia akut dapat mempengaruhi jalan perkembangan sel limfoid akut atau jalur perkembangan sel mieolid akut [7]. Leukemia akut dapat dibagi menjadi:

1. Leukemia Limfositik Akut (LLA)

LLA adalah keganasan klonal dari sel-sel precursor limfoid. Lebih dari 80% kasus, sel-sel ganas berasal dari limfosit B dan sisanya merupakan Leukemia sel T . LLA terjadi pada 80% kasus Leukemia akut anak-anak. Insidensi puncak LLA adalah pada umur 3-7 tahun. LLA juga dapat tampak pada orang dewasa, menyebabkan 20% Leukemia akut dewasa. Tanpa pengobatan rata-rata hidup penderita LLA 3-6 bulan. Dengan pengobatan, 50% penderita LLA yang berumur antara 2-10 tahun rata-rata hidup 2-10 tahun.

2. Leukemia Granulostik/Mielositik Akut (LGA/LMA)

LGA/LMA adalah suatu penyakit yang ditandai dengan transformasi neoplastik dan gangguan differensiasi sel-sel progenitor dari sel myeloid. Pada LGA/LMA terjadi proliferasi dari salah satu unsur sel yang memproduksi sel darah yang ganas. Sel yang ganas tersebut menginfiltrasi sumsum tulang dengan menyebabkan kegagalan fungsi tulang normal dalam proses hematopoetik normal. Dengan pengobatan angka remisi (waktu berkurangnya gejala penyakit) penderita LGA/LMA mencapai 50- 75%, tetapi angka rata-rata hidup masih 2 tahun dan yang dapat hidup lebih dari 5 tahun hanya 10%. Pada saat ini 50% anak-anak dan kira-kira 35% orang dewasa muda disembuhkan dengan kemoterapi intensif. Jika tidak ada pengobatan, penderita LGA/LMA meninggal kira-kira 3-6 bulan.

b. Leukimia kronik [5]

Leukemia kronik ditandai dengan keberadaan jumlah leukosit darah tepi yang sangat tinggi. Sel-sel ini adalah sel matur. Leukemia kronik biasanya memiliki awitan samar dan perkembangannya yang lambat. Sebagian pasien mengalami perkembangan yang lambat dan pembesaran organ yang infiltrasi oleh sel-sel Leukemia.

3. Penjelasan metodologi riset

Data ekspresi gen dari data kanker leukemia dinyatakan dalam matriks W_{MN} dengan entri $w_{ij} \geq 0$ yang merepresentasikan aktifitas gen ke- i pada sampel ke- j . Data ekspresi gen ini dapat dinyatakan menjadi graf bipartite terboboti tak negatif [6]. Titik dapat dikelompokkan ke dalam dua grup, yakni gen dan sampel, dan sisi terboboti antara keduanya ada jika pasangan titik terdapat pada grup yang berbeda (grup gen dan grup sampel). Grup gen dibagi menjadi beberapa kelompok, misalkan A, B, C, \dots . Grup sampel juga dibagi menjadi beberapa kelompok, misalkan $\hat{A}, \hat{B}, \hat{C}, \dots$. Gen di grup A cenderung aktif di sampel \hat{A} dan tidak aktif di sampel lain. Demikian juga gen di grup B aktif di grup \hat{B} dan tidak aktif di grup lain. Secara biologi, gen yang bersama dilibatkan pada fungsi tertentu akan aktif pada himpunan sampel tertentu yang memiliki kesamaan, sebagai contoh memiliki kesamaan penyakit tertentu.

Misalkan $p \in \mathbb{R}^M$ dan $q \in \mathbb{R}^N$ secara berturut-turut merupakan vektor indikator apakah suatu gen dan suatu sampel dikelompokkan pada grup

tertentu. Sampel/gen akan digolongkan menjadi kelompok-kelompok sedemikian sehingga jumlah w_{ij} terboboti yang tidak matching antara gen dan sampelnya menjadi minimum. Sehingga objektifnya adalah akan dicari p_i dan q_j yang meminimumkan resiko ketidaktepatan menempatkan gen i dan sampel j pada grup tertentu. Sehingga masalah pengklasteran dapat didekati dengan masalah minimisasi sebagai berikut:

$$\min \sum_{i=1}^M \sum_{j=1}^N (p_i - q_j)^2 w_{ij} \quad (1)$$

Kemudian dihitung jumlah total ekspresi bobot untuk setiap sampel dari gen tertentu, sebagai contoh gen i , dinotasikan $d_{gen(i)} = \sum_{k=1}^N w_{ik}$, dan jumlah total ekspresi bobot untuk setiap gen dari sampel tertentu, sebagai contoh sampel j , dinotasikan $d_{sampel(j)} = \sum_{k=1}^M w_{kj}$. Dalam bentuk matriks $W_{M \times N}$, $d_{gen(i)}$ adalah jumlah elemen dari baris i dan $d_{sampel(j)}$ adalah jumlah elemen dari kolom ke- j . Dari masing-masing jumlah total elemen baris, dibentuk vektor d_{gen} berukuran M dan vektor d_{sampel} berukuran N yang dibentuk dari masing-masing jumlah total elemen kolom. Dari vektor ini dibentuk matriks diagonal dengan elemen diagonal adalah elemen dari d_{gen} dan d_{sampel} .

$$D_{gen} = \text{diag}(d_{gen}) \in \mathbb{R}^{M \times M} \text{ dan } D_{sampel} = \text{diag}(d_{sampel}) \in \mathbb{R}^{N \times N}$$

dengan asumsi D_{gen} dan D_{sampel} matriks nonsingular.

Dengan notasi tersebut, masalah minimisasi (1) dapat ditulis

$$\min p^T D_{gen} p + q^T D_{sampel} q - 2p^T W q$$

Sehingga dapat dinyatakan dalam masalah maksimisasi:

$$\max p^T W q$$

Supaya gen dan sampel tidak hanya digolongkan ke dalam satu kelompok saja dan level ekspresi terbagi seimbang, maka ditambahkan kendala sebagai berikut:

$$\sum_{i=1}^M p_i d_{gen(i)} \approx 0 \text{ dan } \sum_{j=1}^N q_j d_{sampel(j)} \approx 0$$

Pada penelitian ini ditulis dalam bentuk perkalian vektor:

$$p^T d_{gen} = q^T d_{sampel} = 0$$

Untuk menghindari solusi trivial $p = 0$ dan $q = 0$, ditambahkan kendala normalisasi:

$$\sum_{i=1}^M p_i^2 d_{gen(i)} = 1 \text{ dan } \sum_{i=1}^N q_i^2 d_{sampel(j)} = 1 \quad (2)$$

Persamaan (2) merupakan kendala dimana jika terdapat $d_{gen(i)}$ dan $d_{sampel(j)}$ yang relatif besar (over expressed) maka akan menghasilkan p_i dan q_j yang relatif kecil, mendekati nol. Persamaan (2) dapat ditulis:

$$p^T D_{gen} p = 1 \quad (3)$$

$$q^T D_{sampel} q = 1 \quad (4)$$

Misalkan dinotasikan $D_{gen}^{\frac{1}{2}}$ adalah matriks diagonal dengan elemen diagonal $\sqrt{d_{gen(i)}}$ dan $D_{sampel}^{\frac{1}{2}}$ adalah matriks diagonal dengan elemen diagonal $\sqrt{d_{sampel(j)}}$. Maka persamaan (3) dan (4) dapat ditulis

$$\begin{aligned} \|D_{gen}^{\frac{1}{2}} p\|_2 &= 1 \\ \|D_{sampel}^{\frac{1}{2}} q\|_2 &= 1 \end{aligned}$$

Sehingga masalah yang akan diselesaikan adalah berupa maksimisasi sebagai berikut:

$$\begin{aligned} \max \{ p^T W q : p \in \mathbb{R}^M, q \in \mathbb{R}^N, p^T d_{gen} = q^T d_{sampel} = 0, \\ \|D_{gen}^{\frac{1}{2}} p\|_2 = \|D_{sampel}^{\frac{1}{2}} q\|_2 = 1 \} \end{aligned}$$

Untuk implementasi pada data, akan dilihat bagaimana SVD dapat mengelompokkan data *microarray*. Data *microarray* yang digunakan disini adalah data pasien Leukemia yang digunakan oleh [4] dimana akan dikelompokkan pasien yang termasuk acute lymphoblastic leukemia/leukemia limfositik akut (ALL) dan acute myeloid leukemia/leukemia mielositik akut (AML). Untuk kepentingan implementasi program, pada penelitian ini digunakan program matlab sehingga data dapat dibaca dengan lebih mudah melalui visualisasi.

4. Hasil diskusi

Untuk aplikasi algoritma, digunakan data *microarray* yang dipublikasikan pada [4]. Data terdiri dari 38 data gen pada jaringan tulang belakang pasien dengan ekspresi gen sebanyak 5000 gen. Dua puluh tujuh didiagnosa menderita acute lymphoblastic leukemia (ALL) dan 11 lainnya didiagnosa menderita acute myeloid leukemia (AML). Berikut daftar pasien berdasarkan data pada [4]:

Tabel 1: Diagnosa penyakit pada 38 pasien

NS	Diagnosa	NS	Diagnosa	NS	Diagnosa	NS	Diagnosa
1	<i>ALL_B</i>	11	<i>ALL_B</i>	21	<i>ALL_T</i>	31	<i>AML</i>
2	<i>ALL_B</i>	12	<i>ALL_B</i>	22	<i>ALL_T</i>	32	<i>AML</i>
3	<i>ALL_B</i>	13	<i>ALL_B</i>	23	<i>ALL_T</i>	33	<i>AML</i>
4	<i>ALL_B</i>	14	<i>ALL_B</i>	24	<i>ALL_T</i>	34	<i>AML</i>
5	<i>ALL_B</i>	15	<i>ALL_B</i>	25	<i>ALL_T</i>	35	<i>AML</i>
6	<i>ALL_B</i>	16	<i>ALL_B</i>	26	<i>ALL_T</i>	36	<i>AML</i>
7	<i>ALL_B</i>	17	<i>ALL_B</i>	27	<i>ALL_T</i>	37	<i>AML</i>
8	<i>ALL_B</i>	18	<i>ALL_B</i>	28	<i>AML</i>	38	<i>AML</i>
9	<i>ALL_B</i>	19	<i>ALL_B</i>	29	<i>AML</i>		
10	<i>ALL_B</i>	20	<i>ALL_T</i>	30	<i>AML</i>		

NS: Nomor sampel.

Dari data awal, pertama kolom data diacak dengan tetap mengingat nomor sampel awal. Berikut penomoran baru yang digunakan:

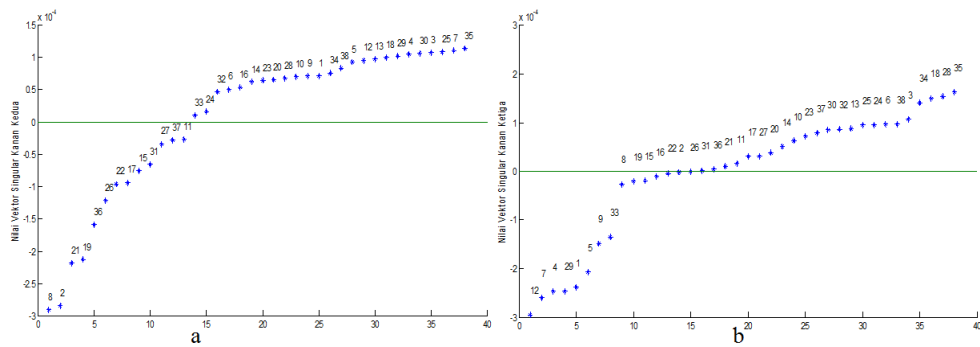
Tabel 2: Diagnosa penyakit pada 38 pasien

NS	1	2	3	4	5	6	7	8	9	10	11	12	13
NB	38	10	23	20	32	27	18	3	30	16	24	28	35
NS	14	15	16	17	18	19	20	21	22	23	24	25	26
NB	25	6	13	37	34	14	33	4	12	29	1	5	9
NS	27	28	29	30	31	32	33	34	35	36	37	38	
NB	7	22	11	8	2	26	17	36	19	15	31	21	

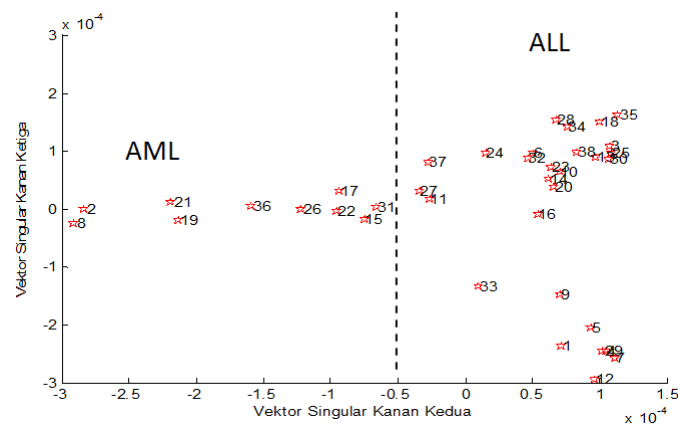
NS: Nomor sampel, NB: Nomor baru.

Berikut pada Gambar 1(a) diplot nilai $D_{smpel}^{-\frac{1}{2}}v^{(2)}$ dengan $v^{(2)}$ merupakan vektor singular kanan kedua dari matriks $D_{gen}^{-\frac{1}{2}}WD_{smpel}^{-\frac{1}{2}}$ setelah diurutkan dari yang terkecil ke yang terbesar. Angka yang terletak di atas titik (*) menunjukkan nomor sampel baru hasil pengacakan. Dari gambar tersebut dapat terlihat adanya dua pengelompokan sampel, yakni yang memiliki nilai vektor singular kanan yang positif dan negatif. Nomor sampel {8, 2, 21, 19, 36, 26, 22, 17, 15, 31, 27, 37, 11} memiliki nilai singular vektor kanan yang kedua yang negatif. Hal ini bersesuaian dengan nomor sampel yang didiagnosa memiliki penyakit AML(lihat tabel 1 dan 2), kecuali nomor sampel 27 dan 37, yang bersesuaian dengan nomor sampel awal 6 dan 17 sebelumnya didiagnosa memiliki penyakit ALL-B. Ketidaksesuaian ini bisa disebabkan karena ketidakkonsistenan data, seperti pembulatan atau pada saat pemilihan gen pada data yang tidak melibatkan semua gen (dalam contoh ini hanya 5000 gen). Ketidaksesuaian ini juga bisa disebabkan kesalahan pada diagnosa sehingga diperlukan metode lebih lanjut untuk menelitinya. Sedangkan sisanya memiliki nilai vektor singular kanan kedua yang positif. Hal ini bersesuaian dengan nomor sampel yang didiagnosa memiliki penyakit ALL. Dari Gambar 1(a) ini terlihat bahwa vektor singular kanan kedua dapat mengelompokkan pasien dengan diagnosa penyakit AML dan ALL. Berikut pada Gambar 1(b) diplot nilai $D_{smpel}^{-\frac{1}{2}}v^{(3)}$ dengan $v^{(3)}$ merupakan vektor singular kanan ketiga dari matriks $D_{gen}^{-\frac{1}{2}}WD_{smpel}^{-\frac{1}{2}}$ setelah diurutkan dari yang terkecil ke yang terbesar.

Dari Gambar 1(b) dapat terlihat adanya pengelompokan sampel. Sampel dengan nomor 12, 7, 4, 29, 1, 5, 9, dan 33 tergolong kedalam satu kelompok yang sama, sebut sebagai kelompok A dan nomor sampel yang lain (8, 19, 15, 16, 22, 2, 26, 31, 36, 21, 11, 17, 27, 20, 14, 10, 23, 37, 30, 32, 13, 25, 24, 6, 38, 3, 34, 18, 28 dan 36) ke dalam kelompok lain, sebut kelompok B. Sampel yang termasuk kelompok A bersesuaian dengan sampel yang didiagnosa memiliki penyakit ALL-T, sedangkan sampel yang termasuk kelompok B didiagnosa memiliki penyakit ALL-B atau memiliki penyakit AML (lihat Tabel 1 dan 2). Dari gambar ini dapat terlihat bahwa nilai vektor singular kanan ketiga dapat mengelompokkan secara nyata sub penyakit ALL, yakni ALL-T dari ALL-B dan AML. Jika diplot antara nilai



Gambar 1: Nilai dari vektor singular kanan kedua (a) dan ketiga (b) setelah pengurutan



Gambar 2: Pengelompokan sampel berdasarkan vektor singular kanan kedua dan ketiga

vektor singular kanan kedua dan ketiga maka akan menghasilkan pengelompokan seperti pada Gambar 2. Pada gambar tersebut dapat terlihat adanya pengelompokan sampel pasien menjadi tiga kelompok. Pada perhitungan ini, nilai singular kesatu sampai keempat yang didapatkan secara berturut-turut adalah 1, 0.38, 0.296 dan 0.234. Adapun persentase kontribusinya jika digunakan singular kedua dan ketiga adalah 13.7. Dengan nilai kontribusi tersebut dapat dikelompokkan data sampel dengan penyakit AML, dan ALL beserta sub tipe penyakit ALL, yakni ALL-T dan ALL-B meski

pengelompokkannya tidak secara nyata, yakni jarak antar kelompok tidak terpisah jauh.

5. Penutup

Berdasarkan hasil dan pembahasan yang telah dijelaskan sebelumnya, maka dapat diambil beberapa kesimpulan bahwa dekomposisi nilai singular dapat digunakan untuk mengelompokkan data dengan ukuran relatif besar dalam hal ini data pasien kanker leukemia. Data *microarray* dengan entri seluruhnya tak negatif dapat dikelompokkan dengan menggunakan nilai singular kedua sebagai hasil dekomposisi nilai singular yang diterapkan pada data dengan menggunakan normalisasi terlebih dahulu. Dan untuk kepentingan visualisasi dapat disertakan juga nilai vektor singular berikutnya (kedua atau ketiga). Dari implementasi pada data, pasien kanker dapat dikelompokkan menjadi penyakit AML, dan ALL beserta sub tipe penyakit ALL, yakni ALL-T dan ALL-B.

Pustaka

- [1] BRAZMA, ALVIS, HELEN P., THOMAS S., MOHAMMADREZA S., *A quick introduction to elements of biology - cells, molecules, genes, functional genomics, microarrays*, European Bioinformatics Institute, on-line tutorial di www.ebi.ac.uk/microarray/biology_intro.html, 2001.
- [2] BRAZMA, ALVIS. AND JAAK VILO, *Gene Expression Data Analysis*, Federation of European Biochemical Societies:480. p. 17-24, 2000.
- [3] BROWN, PATRICK O. AND DAVID BOTSTEINS, . *Exploring the new world of the genome with DNA microarrays*, Nature genetics supplement: 21, p. 33-37, 1999.
- [4] BRUNET, JP., PABLO TAMAYO, T.R. GOLUB, AND J.P. MESIROV, *Metagenes and Molecular Pattern Discovery Using Matrix Factorization.*, Proceedings of The National Academy of Sciences: 101. p. 4164-4169, 2004

-
- [5] CHANDRASOMA, P, TAYLOR, CLIVE R., *Ringkasan Patologi Anatomi*, Edisi ke-2. Jakarta:EGC, 2005.
 - [6] CHEUNG, VIVIAN G , MICHAEL MORLEY, FRANCISCO AGUILAR, ALDO MASSIMI,RAJU KUCHERLAPATI AND GEOFFREY CHILDS, *Making and reading microarrays*, Nature genetics supplement: 21, p. 15-19, 1999
 - [7] RAHMALIA, A, NOVIANTI R., *At a Glance Medicine*, Jakarta: Erlangga., 2006.
 - [8] SETIAWAN, L,*Kapita Selekt Hematologi*, Edisi Keempat. Jakarta: EGC, 2005.